How do I submit flu sequences to GenBank?

NCBI provides a specialized online wizard that will walk you through everything you need to submit influenza A, B or C virus sequences. The wizard has built-in validation steps and can decrease processing time to a matter of hours!

To submit, go to
submit.ncbi.nlm.nih.gov/subs/genbank/

Required Files

☼ FASTA-formatted plain text file of your sequences

☼ Tab-delimited table in a plain-text file with the source information

For single-sequence submissions or smaller sequence sets (<10 sequences), you can add the source information directly in online forms instead of a table. We will add this for larger sequence sets in the future.

```
>XX-5669
ATGAAGACTATCATTGCTTTGAGCTACATTCTATGTCTGGTTTTCGCTCAAAAAATTCCT
GGAAATGACAATAGCACGGCAACGCTGTGCCTTGGGCACCATGCAGTACCAAACGGAACG
ATAGTGAAAACAATCACAAATGACCGAATTGAAGTTACTAATGCTACTGAGTTGGTTCAG
AATTCCTCAATAGGTGAAATATGCGACAGTCCTCATCAGATCCTTGATGGAGAGAACTGC
ACACTAATAGATGCTCTATTGGGAGACCCTCAGTGTGATGGCTTTCAAAATAAGAAATGG
GACCTTTTTGTTGAACGAAGCAAAGCCTACAGCAACTGTTACCCTTATGATGTGCCGGAT
TATGCCTCCCTTAGGTCACTAGTTGCCTCATCCGGCACACTGGAGTTTAACAATGAAAGC
TTCAATTGGACTGGAGTCACTCAAAACGGAACAAGTTCTGCTTGCATAAGGGGATCTAGT
AGTAGTTTCTTTAGTAGATTAAATTGGTTGACCCACTTAAACTACACATATCCAGCATTG
AACGTGACTATGCCAAACAAGGAACAATTTGACAAATTGTACATTTGGGGGGTTCACCAC
```

Sequence_ID	Isolate	Collection_Date	Host	Country	Isolation_source	Serotype	Note
XX-5669	5669	28-Dec-2016	Homo sapiens; gender F; age 92	USA: California	nasal wash	H3N2	passage: PMK1
XX-5685	5685	24-Dec-2016	Homo sapiens; gender F; age 36 mos	Japan	nasal wash	H3N2	passage: PMK1
XX-5689	5689	27-Dec-2016	Homo sapiens; gender F; age 46	USA: New York, Albany	nasal wash	H3N2	passage: PMK1
XX-5699	5699	22-Dec-2016	Homo sapiens; gender F; age 38	South Korea	nasal wash	H3N2	passage: PMK1

Sequence_ID's must be unique within the set and may not contain spaces.

Colection date in the format "DD-Mmm-YYYY", "Mmm-YYYY", or "YYYY" if that's all that is known.

Host can be provided as a common name or a scientific name. More detailed information about the host can be provided after a semicolon, with subsequent fields separated by semicolons. If the sample was not collected directly from a host, use the term "environment."

Geographic location where collected. Must have at least country, but more detailed information can be submitted in the format "country: state, city" or "country: region". The final field will be used when we generate the strain, for example, if the location is "USA: Maryland, Bethesda", the strain will contain "Bethesda."

INSDC country list
ncbi.nlm.nih.gov/genbank/collab/country/

Isolation source describes the host tissue where the sample was collected, or the environment where the sample was collected. Use "missing" if the information was not collected. Isolation source is required for sequences with a host of "environment."
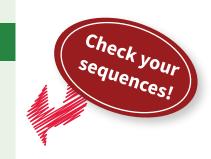
Serotype is required for Influenza A viruses. It must be in format HxNx, Hx, Nx, or mixed; where x is a number.

Note is optional, but often is used for passage details. Individual fields should be separated by semicolons, for example "host did not have symptoms; passage details: 2 passages in MDCK."

IVR Annotation Tool

ncbi.nlm.nih.gov/genomes/FLU/annotation/

Submitted sequences will be annotated using the IVR Annotation Tool. We **strongly** encourage you to check your sequences using this tool before submission. Sequences with errors will delay processing. Find error codes and possible solutions at
ncbi.nlm.nih.gov/sites/genbank/sequencecheck/virus/

Check your sequences!

NIH | **U.S. National Library of Medicine**
National Center for Biotechnology Information

Virus Strain Name

When you submit through the GenBank influenza submission wizard, the strain will be built from the metadata and isolate information you provide. Influenza virus isolate and strain names can reflect special history of a virus.

Sample Strain

A/Maryland/c_beth-627/2009

type | geographic location | isolate name | year

How do I report special isolate names?

Mouse-adapted

Add "-MA" after the isolate string.

✓ **CAN** be submitted through the online wizard

Sample

A/Belgium/145-MA/2009

mouse-adapted

Reverse genetics

The isolate name is

(parental virus isolate)-(institutional code)_RG(clone number)

• Year is the same as the parental virus

✓ **CAN** be submitted through the online wizard

Sample

A/Vietnam/1203-CIP045_RG225/2004

parental virus isolate | institutional code | RG(clone #)

This sample indicates that the virus is a reverse-genetics modification of A/Vietnam/1203/2004.

Reassortant

Use "reassortant" in place of the isolation country, and parental strain information instead of isolate and year.

Sample

A/reassortant/IVR-148(Brisbane/59/2207 x Texas/1/1977)

ressortant instead of country | parental strain information instead of isolate and year

• The parental strains should be ordered by segment, not year (i.e., the H first, the N second, the M third if necessary)

⊘ **CANNOT** currently be submitted through the online flu wizard. Please use the BankIt submission tool.

Looking for more information?

A description of the submission processes and details on the required files are provided at **submit.ncbi.nlm.nih.gov/genbank/help/**

We need your help!

We are hoping to provide more online wizards like this for other viruses in the future – but they depend on help from the research community to provide a good reference set, and widely-accepted gene and protein names. If you think you can help, please contact us at **info@ncbi.nlm.mih.gov** and let us know which virus(es) you would like to help with.

NIH U.S. National Library of Medicine
National Center for Biotechnology Information